On Generalizing the *AMG* Framework

Robert D. Falgout and **Panayot Vassilevski**

Center for Applied Scientific Computing
Lawrence Livermore National Laboratory

Germany
June, 2003

Outline

- *AMG / AMGe* framework background
- New Measures and Convergence Theory
- Building Interpolation
- Compatible Relaxation
- Examples
- Conclusions and future directions

AMG / AMGe Framework

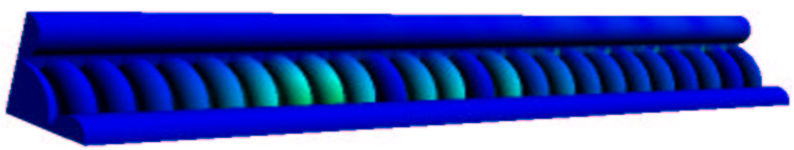
- ***AMGe* heuristic is based on multigrid theory:** *interpolation must reproduce a mode up to the same accuracy as the size of the associated eigenvalue*
- **Bound a measure (weak approximation property):**

$$\|A\| \frac{\langle (I-Q)e, (I-Q)e \rangle}{\langle Ae, e \rangle}; \quad Q = P\begin{bmatrix} 0 & I \end{bmatrix}$$

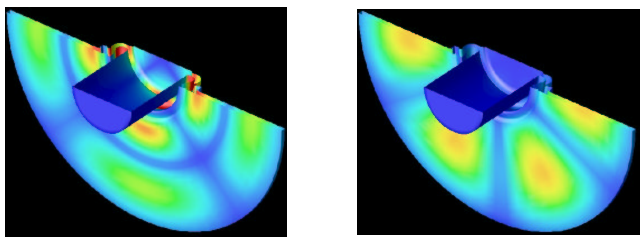
- **Localize the measure to build *AMGe* components**
- **Several variants developed: *E-Free, Spectral***
- **Based on pointwise relaxation**
- **Assumes coarse grid is a subset of fine grid**

We are generalizing our *AMG* framework to address new problem classes

- **Maxwell and Helmholtz problems have huge near null spaces and require more than pointwise smoothing to achieve optimality in multigrid**



Model of a section of the Next Linear Collider structure



Resonant frequencies in a Helmholtz Application

- **Our new theory allows for any type of smoother, and also works for a variety of coarsening approaches (e.g., vertex-based, cell-based, agglomeration)**
- **Paper submitted**

Preliminaries…

- **Consider solving the linear system**

$$Au = f$$

- **Consider smoothers of the form**

$$u_{k+1} = u_k + M^{-1} r_k$$

where we assume that $(M+M^T-A)$ **is SPD (necessary & sufficient condition for convergence)**

- **Note:** M **may be symmetric or nonsymmetric**

- **Smoother error propagation**

$$e_{k+1} = (I - M^{-1}A)\, e_k$$

Preliminaries continued…

- **Let** $P : \Re^{n_c} \rightarrow \Re^{n}$ **be interpolation (prolongation)**

- **Let** $R : \Re^{n} \rightarrow \Re^{n_c}$ **be some "restriction" operator**
  - Note that R is not the MG restriction operator
  - The form of R will be important later

- **Define** $Q : \Re^{n} \rightarrow \Re^{n}$ **to be a projection onto** $\text{range}(P)$**; hence** $Q = PR$ **such that** $RP = I$

Two new measures

- **First measure:**

$$\mu(Q,e) = \frac{\left\langle M(M+M^T-A)^{-1}M^T(I-Q)e,\,(I-Q)e\right\rangle}{\langle Ae,\,e\rangle}$$

- **Second measure:** Define $\sigma(M) \equiv \frac{1}{2}(M+M^T)$, then

$$\mu_\sigma(Q,e) = \frac{\left\langle \sigma(M)\,(I-Q)e,\,(I-Q)e\right\rangle}{\langle Ae,\,e\rangle}$$

- **Measure μ_σ is the analogue to the *AMGe* measure**

First measure and *MG* convergence

- **Theorem:** Assume that the following holds for some constant K:

$$\mu(Q,e) \le K \quad \forall e \in \Re^n \setminus \{0\}$$

Then, 2-level *MG* converges uniformly:

$$\left\| (I - M^{-1}A)(I - Q_A)\, e \right\|_A \le \left(1 - \frac{1}{K} \right)^{1/2} \| e \|_A$$

Here, $Q_A = P(P^TAP)^{-1}P^TA$ is the A-orthogonal projector onto $\mathrm{range}(P)$

- **As in *AMGe*, we could try to directly localize this new measure to help us build *AMG* algorithms**
- **But, we will take a different approach**

Second measure and MG convergence

- **Bounding μ_σ also implies uniform convergence…**
- **Lemma:** Assume that $(M+M^T-A)$ is SPD. Then,

$$\mu(Q,e) \le \frac{\Delta^2}{2-\omega}\mu_\sigma(Q,e)$$

where $\Delta \ge 1$ measures the deviation of M from $\sigma(M)$

$$\langle Mv,w\rangle \le \Delta \langle \sigma(M)v,v\rangle^{1/2}\langle \sigma(M)w,w\rangle^{1/2}$$

and where $0 < \omega \equiv \lambda_{max}(\sigma(M)^{-1}A) < 2$.

- **Must insure "good" constants**
  - in particular, $\omega \ll 2$

General notions of *C-pts* & *F-pts*

- **Recall the projection** $Q=PR$**, with** $RP=I$
- **We now fix** R **so that it does not depend on** P
  - Defines the coarse-grid variables, $u_c = Ru$
  - Recall that $R=[\ 0, I\ ]$ $(P^T=[\ W^T, I\ ]^T)$ for *AMGe*; i.e., the coarse-grid variables were a subset of the fine grid
  - *C-pt* analogue
- **Define** $S : \Re^{n_s} \rightarrow \Re^n$ **s.t.** $n_s = n - n_c$ **and** $RS = 0$
  - Think of $\mathrm{range}(S)$ as the "smoother space", i.e., the space on which the smoother must be effective
  - Note that S is not unique
  - *F-pt* analogue
- S **and** R^T **define an orthogonal decomposition of** $\Re^n$**; any vector** e **can be written as** $e = Se_s + R^Te_c$

The Min-max Problem

- **Consider the following base measure, where X is any SPD matrix:**

$$\mu_X(Q, e) \equiv \frac{\langle X(I-Q)e, (I-Q)e \rangle}{\langle Ae, e \rangle}$$

- **Theorem:** Define

$$\mu_X^* \equiv \min_P \max_{e \neq 0} \mu_X(PR, e)$$

The *arg min* satisfies $S^T A P_* = 0$ and the minimum is

$$\mu_X^* = \lambda_{\min}^{-1}\left((S^T X S)^{-1} (S^T A S) \right)$$

- **We will call P_* the optimal interpolation operator**

The Min-max Problem… and AMGe

- **The optimal interpolation has the general form:**

$$P_* = \begin{bmatrix} S & R^T \end{bmatrix} \begin{bmatrix} -(S^T A S)^{-1}(S^T A R^T) \\ I \end{bmatrix}$$

- **For AMGe, the coarse-grid variables are a subset of the fine grid:**

$$R = \begin{bmatrix} 0 & I \end{bmatrix}; \quad P = \begin{bmatrix} W \\ I \end{bmatrix}; \quad S = \begin{bmatrix} I \\ 0 \end{bmatrix}$$

Hence,

$$P_* = \begin{bmatrix} -A_{ff}^{-1} A_{fc} \\ I \end{bmatrix}, \quad \mu_X^* = \frac{\|A\|}{\lambda_{\min}(A_{ff})}$$

The Min-max Problem… Spectral AMGe and Smoothed Aggregation (SA)

- **For Spectral AMGe and SA, the coarse-grid variables are coefficients of basis functions:**

$$R^T = \left[p_1, \ldots, p_c \right], \quad RP = I, \quad S = \left[p_{c+1}, \ldots, p_n \right]$$

where the p_i **are orthonormal eigenvectors of** A **with eigenvalues** $\lambda_1 \le \ldots \le \lambda_n$**. Hence,**

$$P_* = R^T, \quad \mu_X^* = \frac{\lambda_n}{\lambda_{c+1}}$$

- **The optimal interpolation can also be viewed as a "smoothed" tentative prolongator**

$$P_* = (I - S(S^T A S)^{-1} S^T A) \; R^T$$

The new theory separates construction of coarse-grid correction into two parts

- **The following measures the ability of a given coarse grid Ω_c to represent algebraically smooth error:**

$$\mu^* \equiv \min_P \max_{e \neq 0} \mu(PR, e)$$

- **Theorem:** (1) Assume that $\mu^* \leq K$ for some constant K. (2) Assume that any one of the following holds for $\eta \geq 1$:

$$\langle A\,Qe, Qe \rangle \leq \eta \langle Ae, e \rangle, \quad \forall e$$

$$\langle A\,(I-Q)\,e, (I-Q)\,e \rangle \leq \eta \langle Ae, e \rangle, \quad \forall e$$

$$\left\langle APe_c, Se_s \right\rangle^2 \leq (1-\eta^{-1}) \left\langle APe_c, Pe_c \right\rangle \left\langle ASe_s, Se_s \right\rangle, \quad \forall\, e_c, e_s$$

Then, $\mu(PR, e) \leq \eta K, \forall e$.

- **(1) insures coarse grid quality** – use *CR*
- **(2) insures interpolation quality** – necessary condition that does not depend on relaxation!

CR is an efficient method for measuring the quality of the set of coarse variables

- ***CR* (Brandt, 2000) is a modified relaxation scheme that keeps the coarse-level variables, Ru, invariant**
- **We have defined several variants of *CR*, and shown that fast converging *CR* implies a good coarse grid:**

$$\mu^* \leq \left(\frac{\Delta^2}{2-\omega} \right) \frac{1}{1-\rho_{cr}}$$

- **Hence, *CR* can be used as a tool to efficiently measure the quality of a coarse grid!**
- **General idea:** *If CR is slow to converge, either increase the size of the coarse grid or modify relaxation*
- ***F*-relaxation is a specific instance of *CR***

We can use *CR* to choose the coarse grid

- **To check convergence of *CR*, relax on the equation**

$$A_{ff}\, x = 0$$

  and monitor pointwise convergence to *0*
- ***CR* coarsening algorithm:**

$$\text{Initialize } U = \Omega;\ C = \varnothing;\ F = \Omega - C$$

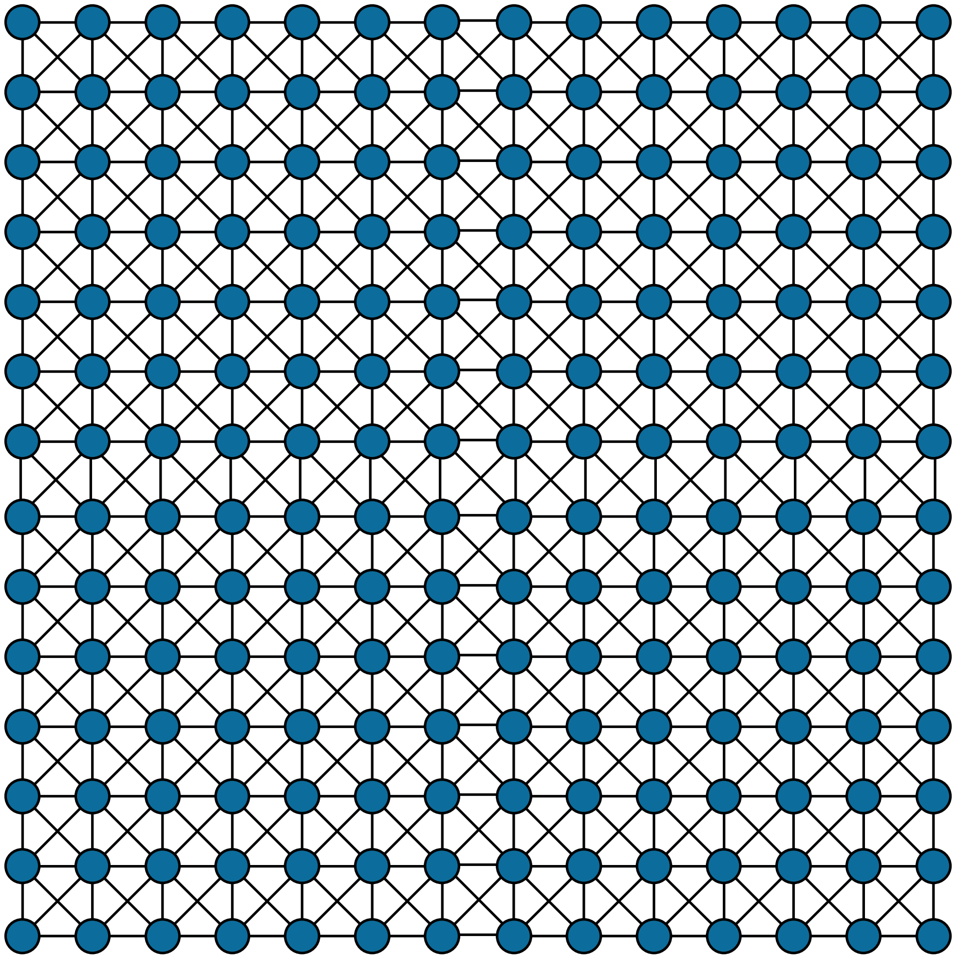
$$\text{While } U \neq \varnothing$$

$$\text{Do } \nu \text{ compatible relaxation sweeps}$$

$$U = \{ i : |x_i^{\nu} / x_i^{\nu - 1}| > \theta \}$$

$$C = C \cup \{ \text{independent set of } U \} ;\ F = \Omega - C$$

Using *CR* to choose the coarse grid

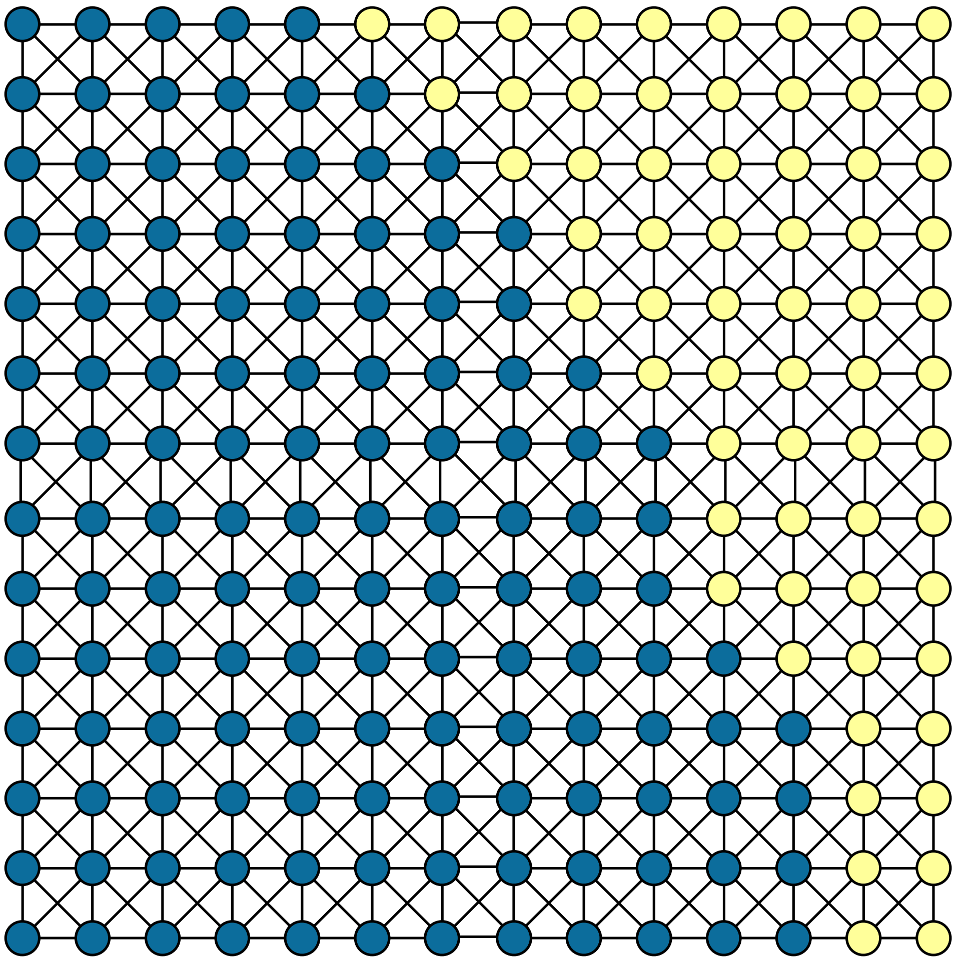


➔ **Initialize U-pts**

➔ **Do CR and redefine U-pts as points slow to converge**

➔ **Select new C-pts as indep. set over U**

Using *CR* to choose the coarse grid

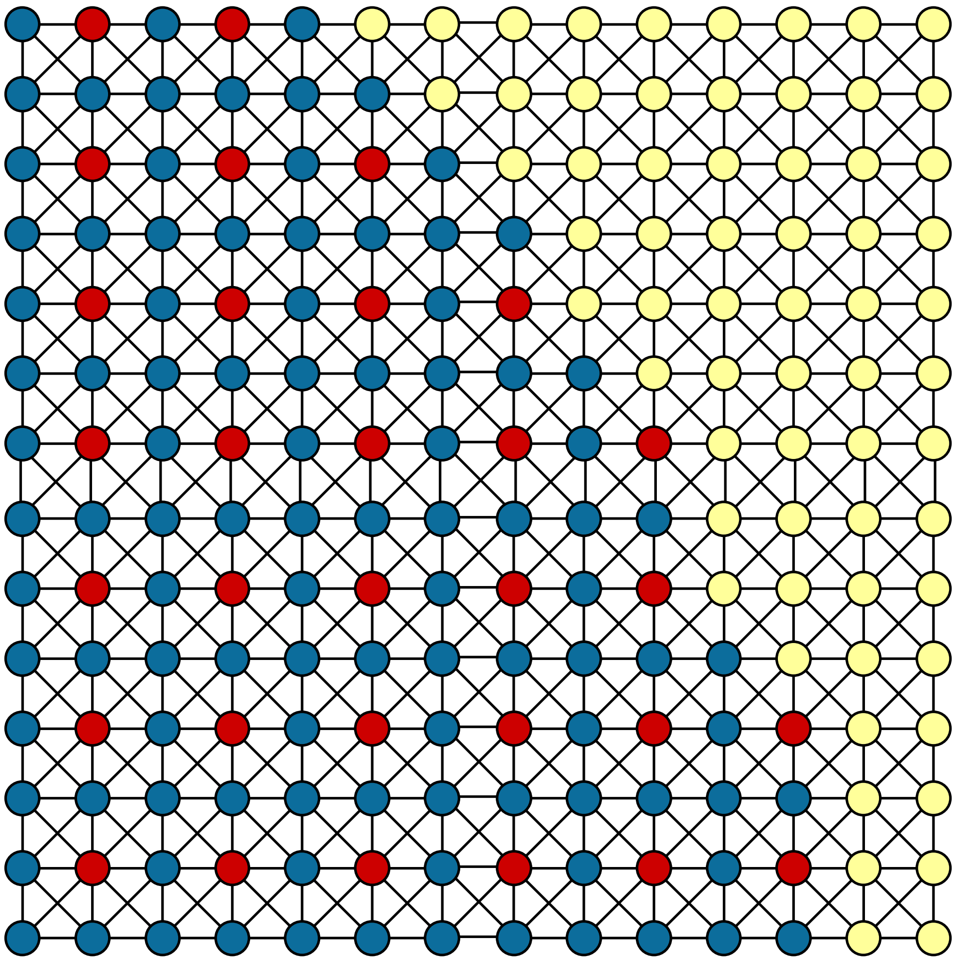


➔ Initialize U-pts

➔ **Do CR and redefine U-pts as points slow to converge**

➔ Select new C-pts as indep. set over U

Using *CR* to choose the coarse grid

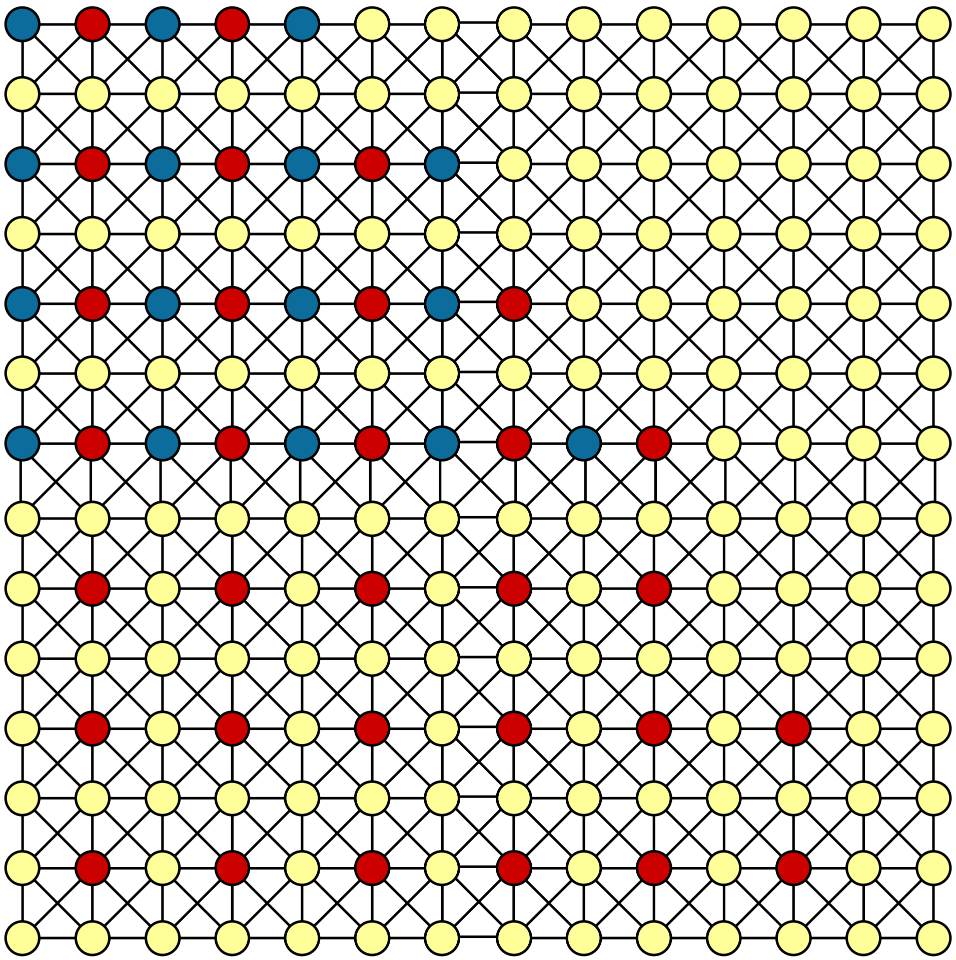


➔ Initialize U-pts

➔ Do CR and redefine U-pts as points slow to converge

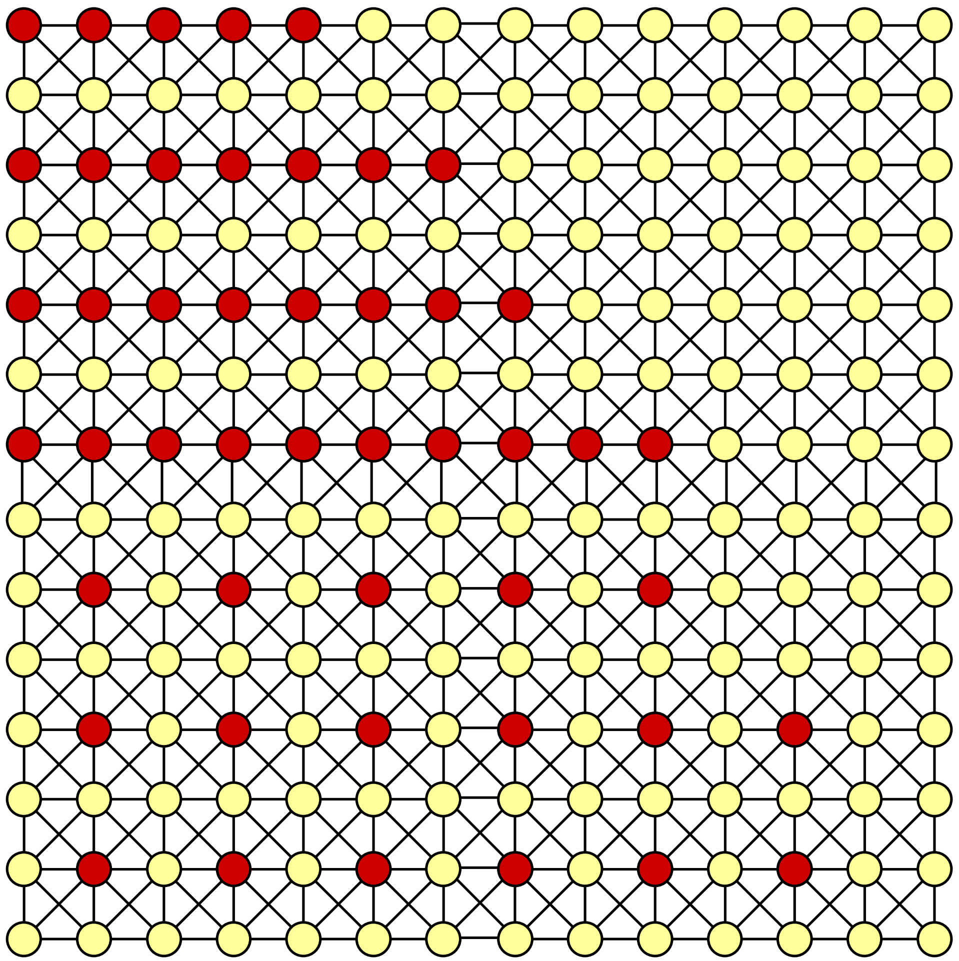
➔ **Select new C-pts as indep. set over U**

Using *CR* to choose the coarse grid



- ➔ Initialize U-pts
- ➔ **Do CR and redefine U-pts as points slow to converge**
- ➔ Select new C-pts as indep. set over U

Using *CR* to choose the coarse grid



➔ Initialize U-pts

➔ Do CR and redefine U-pts as points slow to converge

➔ **Select new C-pts as indep. set over U**

CR based on matrix splittings

$$e_{k+1} = (I - M_s^{-1} A_s)\, e_k;\ M_s = S^T M S;\ A_s = S^T A S$$

- **Theorem:** Assume that $(M + M^T - A)$ is SPD. Then,

$$\mu^* \le \left(\frac{\Delta^2}{2 - \omega} \right) \frac{1}{1 - \rho_s}$$

  where Δ and ω are as before, and $\rho_s = \| (I - M_s^{-1} A_s) \|_{A_s}$.
- **Fast converging *CR* implies good coarse grid**
- **If relaxation is based on a splitting $A = M - N$, then M is explicitly available, and *CR* is probably feasible**

CR based on additive subspace methods

- **Consider the following additive method:**

$$I - M^{-1}A; \;\; M^{-1} = \sum_i I_i (I_i^T A I_i)^{-1} I_i^T$$

where $I_i : \Re^{n_i} \to \Re^n$ **and** $\Re^n = \cup_i \,\mathrm{range}(I_i)$.

- **Define full rank normalized operators** S_i **and** R_i^T **s.t.** $\mathrm{range}(S_i) = \mathrm{range}(I_i^T S)$ **and** $\mathrm{range}(R_i^T) = \mathrm{range}(I_i^T R^T)$
- **The** I_i **must be chosen so that** $R_i S_i = 0$
- **Then an additive *CR* is given by**

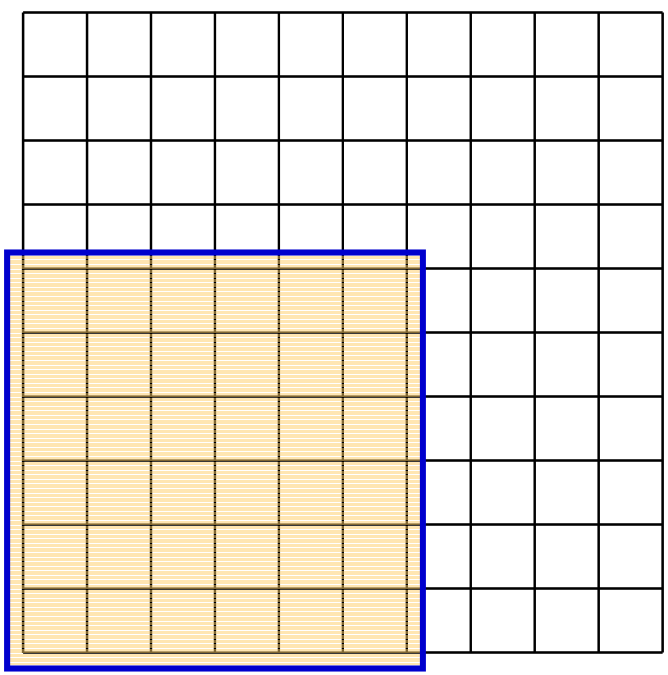
$$I - M_{cr}^{-1} A_s; \;\; M_{cr}^{-1} = \sum_i S^T I_{s,i} (I_{s,i}^T A I_{s,i})^{-1} I_{s,i}^T; \;\; I_{s,i} = I_i S_i$$

- **Same theoretical result as before, but with** $\Delta = 1$

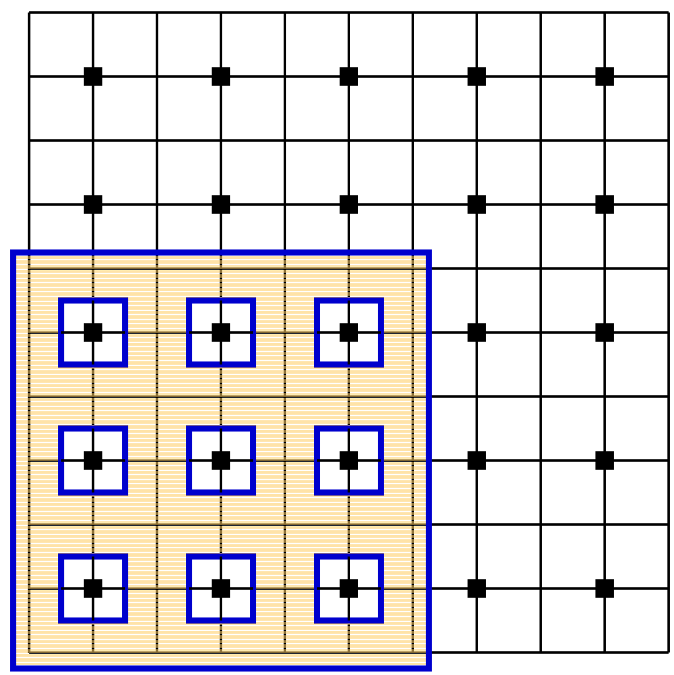
Compatible Additive Schwarz is natural when $R=[\,0, I\,]$

- **Just remove coarse-grid points from subdomains**
- **It is clear that $R_i S_i = 0$ for any choice of I_i**

Additive Schwarz



CR Additive Schwarz



More general form of *CR*

$$e_{k+1} = (I - (S^T M^{-1} S)\, A_s)\, e_k;\; A_s = S^T A S$$

- **Here,** S **must be normalized so that** $S^T S = I$
- **This variant of *CR* is always computable**
- **Theoretical result currently requires SPD smoother,** M**, and involves an additional constant:**

$$\mu^* \le \left(\frac{1}{2-\omega}\right)\left(\frac{1}{1-\gamma^2}\right)\frac{1}{1-\rho_s}$$

where $\gamma \in [0,1)$ **satisfies**

$$\left\langle MSv_s, R^T v_c \right\rangle \le \gamma \left\langle MSv_s, Sv_s \right\rangle^{1/2} \left\langle MR^T v_c, R^T v_c \right\rangle^{1/2};\;\; \forall\, v_s, v_c$$

Another general form of *CR* (due to Brandt and Livne)

$$e_{k+1} = (I - S^T M^{-1} A S)\, e_k = S^T (I - M^{-1} A)\, S e_k$$

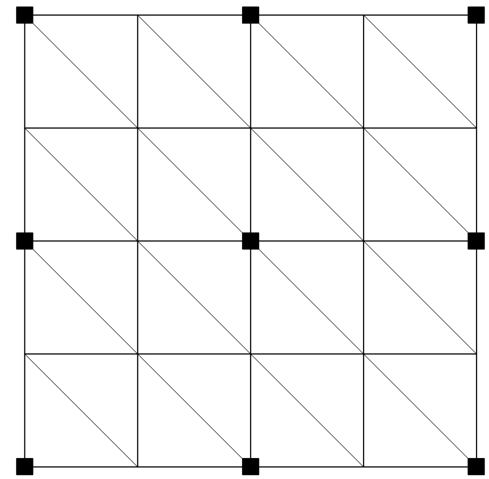
- **As before, S must be normalized so that $S^T S = I$**
- **This variant of *CR* is also always computable**

- **Theoretical result is similar, but weaker:**

$$\mu^* \le \left(\frac{1}{2-\omega} \right) \left(\frac{2}{1-\gamma^2} \right) \frac{1}{(1-\rho_s)^2}$$

Anisotropic Diffusion Example

$$-\varepsilon u_{xx} - u_{yy} = f$$



- **Dirichlet BC's and** $\varepsilon \in (0,1]$
- **Piecewise linear elts on triangles**
- **Standard coarsening, i.e.,** $S = [\, I, 0 \,]^T$

- **The spectrum of the *CR* iteration matrix satisfies**

$$\lambda(\, I - M_S^{-1} A_S) \in \left[-\sqrt{\frac{\varepsilon}{2+\varepsilon}}, \sqrt{\frac{\varepsilon}{2+\varepsilon}} \right]$$

- **Linear interpolation satisfies, with** $\eta = 2$**,**

$$\langle A\, Qe,\, Qe \rangle \leq \eta \, \langle Ae,\, e \rangle, \quad \forall e$$

Anisotropic Diffusion Example – leveraging previous work

- **Consider the AMGe measure**

$$\|A\| \, \|(I-Q)\,e\|^2 \le \eta \, \langle Ae,\, e \rangle$$

- **It is easy to show that** $\eta \ge \|A\| \,/\, \varepsilon$
- **As mentioned earlier, this implies**

$$\langle A\,(I-Q)\,e,\,(I-Q)\,e \rangle \le \eta \, \langle Ae,\, e \rangle, \quad \forall e$$

- **But the AMGe method produces linear interpolation; it is just unable to judge its quality in this setting (i.e., when using line relaxation)**

Conclusions and Future Directions

- **We have developed a more general theoretical framework for AMG methods**
  - Allows for any type of smoother
  - Allows for a variety of coarsening approaches (e.g., vertex-based, cell-based, agglomeration)
- **The theory separates construction of coarse-grid correction into two parts:**
  - Insuring the quality of the coarse grid via *CR*
  - Insuring the quality of interpolation for the given coarse grid (leverages earlier work)
- **We have defined several variants of *CR***
- **Will explore further the use of *CR* in practice**
- **Choosing / modifying smoothers automatically?**

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48.

UCRL-PRES-150807